

aDORe Archive - Overview

Table of contents

1 What is the aDORe Archive?.....	2
2 How does the aDORe Archive work?.....	2
3 Additional Information.....	3

1. What is the aDORe Archive?

The aDORe Archive is a write-once/read-many storage approach for Digital Objects and their constituent datastreams. The approach combines two interconnected file-based storage mechanisms that are made accessible in a protocol-based manner. First, XML-based representations of multiple Digital Objects are concatenated into a single, valid XML file named an XMLtape. The creation of indexes for both the identifier and the creation datetime of the XML-based representation of the Digital Objects, facilitates OAI-PMH-based access. Second, ARC files, as introduced by the Internet Archive, are used to contain the constituent datastreams of the Digital Objects in a concatenated manner. An index for the identifier of the datastream facilitates OpenURL-based access. The interconnection between an XMLtape and its associated ARC file(s) is provided by conveying the identifiers of these ARC files as administrative information in the XMLtape, and by including OpenURL references to constituent datastreams of a Digital Object in the XML-based representation of that Digital Object stored in the XMLtape.

The aDORe Archive allows for the storage of multiple XMLtapes and ARC files through the introduction of OAI-PMH compliant XMLtape and ARCfile registries.

The aDORe Archive Solution provides:

- Storage of compound objects (independent of the choice of complex object format, i.e. MPEG-21 DIDL, METS, ...)
- Two interconnected file-based storage mechanisms:
 - XMLtapes: File storage of XML-based representations of Digital Objects
 - ARC files: File storage of constituent datastreams of Digital Objects
 - An XMLtape is interconnected with one or more ARC files during the ingestion process
- Protocol-based access mechanisms:
 - Each XMLTape is exposed as an autonomous OAI-PMH repository
 - Each ARC file is exposed as an OpenURL Resolver
- Long-term stability::
 - Write once - Read many approach
 - XMLtapes and ARC files remain stable over time, while indexing mechanisms can change as technologies evolve
 - Protocol access remains stable over time as indexing technologies evolve
- Capability to store multiple XMLtapes and ARCfiles

2. How does the aDORe Archive work?

- XMLTape
 - A Digital Object is represented and packaged using a Complex Object format (e.g. MPEG-21 DID), then stored in an XMLtape. Multiple Digital Objects can be stored in an XMLtape.
 - Constituent datastreams of a Digital Object are provided By-Reference using a reference capability that is typically available in complex object formats. The constituent datastreams are stored in an ARC file.
 - An XMLtape is connected with one or more ARC files by including the ARCfile Identifiers in the XMLtape-level administrative section.
 - Mandatory Tape Record Admin metadata (identifier, datestamp) is indexed using specified implementation.
 - An XMLtape becomes available as an OAI-PMH repository.
- ARCFile
 - Constituent datastreams of a Digital Object are stored in an Internet Archive ARC file (URI of an ARC record = Datastream Identifier).
 - ARC Record Admin metadata (Datastream identifier) is indexed and serialized to CDX file format.
 - An ARC files becomes available as an OpenURL Resolver.
- XMLTape Registry
 - An administrative OAI-PMH Repository, which keeps track of the creation and location of XMLtapes.
- ARCFile Registry
 - An administrative OAI-PMH Repository, which keeps track of the creation and location of ARCFiles.
- Archive Accessor
 - An OAI-PMH compliant front-end to the aDORe Archive through which XML-based representations of Digital Objects can be requested.
- ARCFile OpenURL Resolver
 - An OpenURL compliant front-end to the aDORe Archive through which constituent datastreams of Digital Objects can be requested.

Figure 1

[Complete aDORe Archive Diagram - PDF Version](#)

3. Additional Information

Liu, X., Balakireva, L., Hochstenbach, P., van de Sompel, H. (2005, June).

[File-based storage of Digital Objects and constituent datastreams: XMLTapes and Internet Archive ARC files](#)

Liu, X., Balakireva, L., Hochstenbach, P., van de Sompel, H. (2005, September).
[File-based storage of Digital Objects and constituent datastreams: XMLTapes and Internet Archive ARC files \(Presentation\)](#)

van de Sompel, H., Hammond, T., Neylon, E., Weibel, E. (2005, August).
[The "info" URI Scheme for Information Assets with Identifiers in Public Namespaces](#)

Burner, M., Kahle, B (1996, September).
[Arc File Format](#)